Diagonalization

Def: Two $n \times n$ matrices A and B are called **similar** if $A = PBP^{-1}$ for some invertible matrix P

(Note: $A = PBP^{-1} \iff B = P^{-1}AP = RBR^{-1}$ where $R = P^{-1}$)

We care about similar matrices because of the following:

Property: If A and B are similar, then A and B have the same eigenvalues.

Proof: Suppose $A = PBP^{-1}$ and let λ be an eigenvalue of B. Let v be a corresponding eigenvector. Then $Bv = \lambda v$.

Now $A(Pv) = (PBP^{-1})Pv = PB(P^{-1}P)v = PBv = P(\lambda v) = \lambda(Pv)$

Hence Pv is an eigenvector of A corresp to λ. ∎

Def: A is called **diagonalizable** if it is similar to a diagonal matrix.

Ex: $\begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{-1}$ (check)

$A \qquad P \qquad D \qquad P^{-1}$

We'll see that diagonalizability is useful for Markov Chains

How do we determine if a matrix is diagonalizable?

Diagonalizability Theorem

A is diagonalizable if and only if it has n linearly independent eigenvectors.

Moreover, if A is diagonalizable, then $A = PDP^{-1}$,

where $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ and $P = [v_1 \cdots v_n]$,

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A

and $v_1, \ldots, v_n$ are corresponding eigenvectors.

(ie. v_i corresponds to λ_i).

Proof:

($\Rightarrow$) A diagonalizable $\Rightarrow A = PDP^{-1} \Rightarrow AP = PD$

Let $P = [v_1 \cdots v_n]$ and $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. Since P is invertible, $\{v_1, \ldots, v_n\}$ is L.I. Now, multiplying out $AP = PD$ gives $[Av_1 \cdots Av_n] = [\lambda_1 v_1 \cdots \lambda_n v_n]$. Hence $Av_i = \lambda_i v_i \ \forall\ 1 \le i \le n$

$\Rightarrow \{v_1, \ldots, v_n\}$ are n L.I. eigenvectors.

($\Leftarrow$) Let $v_1, \ldots, v_n$ be n L.I. eigenvectors with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Set $P = [v_1 \cdots v_n]$ and $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.

Then $AP = [Av_1 \cdots Av_n] = [\lambda_1 v_1 \cdots \lambda_n v_n] = PD$

$\rightarrow A = PDP^{-1} \Rightarrow$ A is diagonalizable ■

Key Fact: Eigenvectors from distinct eigenspaces are linearly independent

Ex: Diagonalize $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$, if possible

Since A is triangular, its eigenvalues are 1 and 3

We need to find three L.I. eigenvectors.

$E_1(A) = \text{Nul}(A - I) = \text{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}\right\}$ (Check)

$E_3(A) = \text{Nul}(A - 3I) = \text{Span}\left\{\begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}\right\} = \text{Span}\left\{\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right\}$ (Check)

It follows from the key fact that $\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right\}$ is L.I.

Since $\exists$ 3 L.I. evectors, A is diagonalizable and

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\Rightarrow \quad A = PDP^{-1}$$ (Check by computing AP and PD)

Important (for Markov Chains):

Let $A = PDP^{-1}$, where $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

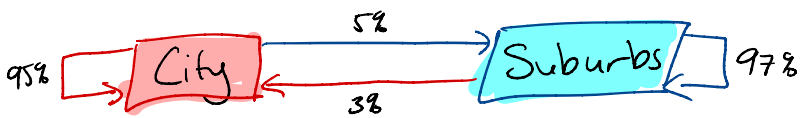
Then $A^k = (PDP^{-1})^k = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = PD^kP^{-1}$

Since $D^k = \begin{bmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{bmatrix}$ is diagonal, A^k is diagonalizable

(and has eigenvalues $\lambda_1^k, \ldots, \lambda_n^k$, which we already knew)

Markov Chain Example, revisited

Recall the example given by



The associated Markov matrix is $M = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$

Diagonalizing M allows us to find a formula for M^k

$$M = \begin{bmatrix} .6 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .92 \end{bmatrix} \begin{bmatrix} .6 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \quad \text{(check)}$$

Hence $M^k = \begin{bmatrix} .6 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .92^k \end{bmatrix} \begin{bmatrix} .6 & -1 \\ 1 & 1 \end{bmatrix}^{-1}$

$$= \begin{bmatrix} .375 + (.625)(.92)^k & .375 - (.375)(.92)^k \\ .625 - (.625)(.92)^k & .625 - (.375)(.92)^k \end{bmatrix}$$

Let $x_0 = \begin{bmatrix} a \\ b \end{bmatrix}$ be <u>any</u> initial state vector (so $a+b=1$). Then

$$x_k = M^k x_0 = \begin{bmatrix} .375 + (.225)(.92)^k & .375 - (.375)(.92)^k \\ .625 - (.225)(.92)^k & .625 - (.375)(.92)^k \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

and as $k \to \infty$, we have

$$x_k \longrightarrow \begin{bmatrix} .375 & .375 \\ .625 & .625 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} .375(a+b) \\ .625(a+b) \end{bmatrix} = \begin{bmatrix} .375 \\ .625 \end{bmatrix}$$

Originally, we solved a system to express x_0 as a L.C. of eigenvectors, and then computed $x_k = M^k x_0$. Here, we used diagonalizability to compute x_k.